

# Estudando Dinâmicas de Grupo Através da Utilização da Análise de Redes Sociais em uma Comunidade Médica

Ruben A. Perorazio<sup>1</sup>, Fabrício F. Faria<sup>1</sup>, Rafael Monclar<sup>2</sup>, Jonice Oliveira<sup>1</sup>, Jano M. de Souza<sup>2</sup>

<sup>1</sup>Programa de Pós-Graduação em Informática – Universidade Federal do Rio de Janeiro (UFRJ)

<sup>2</sup>Programa de Engenharia de Sistemas e Computação – COPPE - Universidade Federal do Rio de Janeiro (UFRJ)

jonice@dcc.ufrj.br, {rrpero, firminodefaria}@ppgi.ufrj.br, {rastumon, jano}@cos.ufrj.br

**Abstract.** *Due to importance of National Institutes of Science and Technology (INCT) for Cancer Control and different groups working together in the same area, it's important that they collaborate among themselves. Therefore, the BRINCA project was created and several scientific networks' analyses were created. These analyses were produced by mining the Curricula Lattes and PubMed. The analyses were used by the INCT's supervision to understand and to improve scientific collaborations.*

**Resumo.** *Dada a importância do INCT para o Controle do Câncer e a existência de diversos grupos trabalhando nessa mesma temática, é importante que eles colaborem entre si. Para isto, foi criado o projeto BRINCA e nele foi conduzida uma série de análises das redes científicas deste INCT. Essas análises foram conduzidas fazendo uso de mineração de dados do Lattes e do PubMed. As análises serviram para que a coordenação deste INCT pudesse realizar ações para entender e melhorar as colaborações científicas.*

## 1. Introdução

O governo brasileiro criou os Institutos Nacionais de Ciência e Tecnologia [INCT, 2012] que visa reunir diferentes pesquisadores em áreas estratégicas. Um desses institutos é o INCT<sup>1</sup> para o Controle do Câncer, que é coordenado pelo INCA<sup>2</sup>. Neste cenário, o projeto BRINCA (Análise e Balanceamento de Redes Sociais Científicas no Controle do Câncer) foi criado. Os principais objetivos do projeto são analisar a forma como os membros do INCT para o Controle do Câncer colaboraram e como o conhecimento científico fluiu entre os diferentes pesquisadores e instituições do grupo.

---

<sup>1</sup> INCT - Institutos Nacionais de Ciência e Tecnologia

<sup>2</sup> INCA - Instituto Nacional de Câncer

## 2. Projeto BRINCA

A seguir, descrevemos detalhes sobre o ambiente criado para a análise da rede social do INCTCC.

### 2.1. Arquitetura

A arquitetura que desenvolvemos para o nosso trabalho engloba as etapas mostradas na Figura 1.

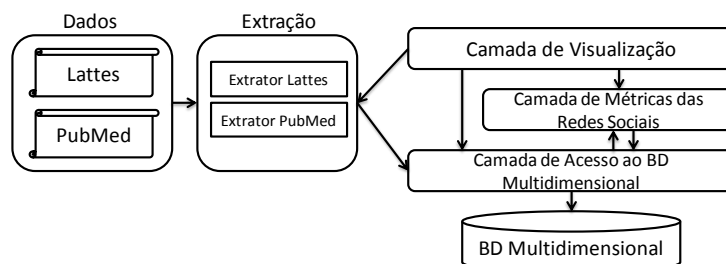


Figura 1. Arquitetura do Ambiente BRINCA

As fontes de dados utilizadas foram obtidas através do Lattes [2012] e do PubMed [2012]. Os currículos Lattes foram capturados através da ferramenta XML-Lattes [Fernandes, 2011]. O processo de extração, tratamento e limpeza foi orquestrado através do Kettle [Pentaho, 2012]. Por fim, com base no Data Warehouse gerado realizamos a visualização através do Gephi [2012] e a análise através do Tableau [2012]. As métricas de redes sociais utilizadas foram calculadas com o Gephi [2012] e armazenadas no DW. Na seção 3 veremos em maiores detalhes tais visualizações e análises. O modelo multidimensional é descrito na Figura 2.

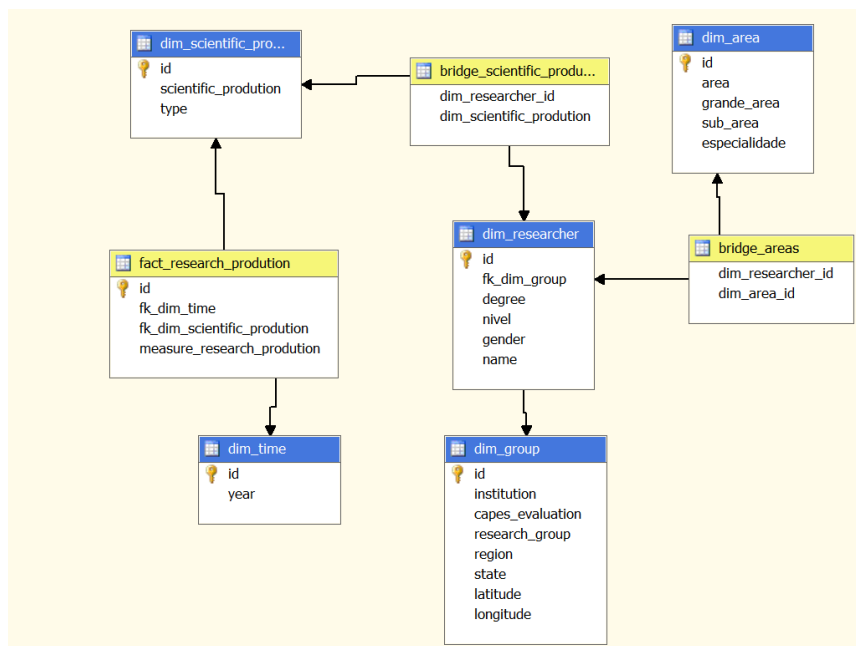


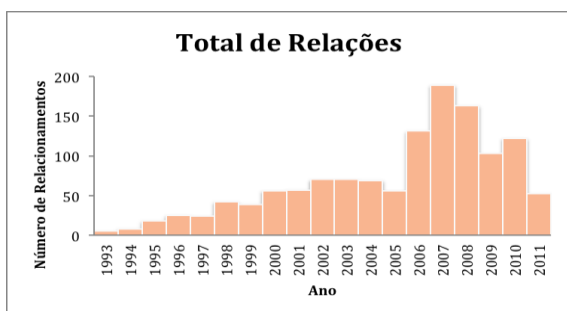
Figura 2. Modelo Multidimensional

### 3. Resultados Atuais

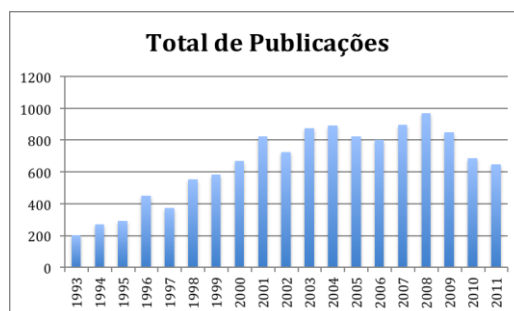
No projeto do INCT para o Controle do Câncer (INCTCC), existem **temas** de pesquisa e dentro de cada tema existem **subprojetos** [INCTCC 2010]. Ligados a esses subprojetos temos diversos pesquisadores. Para analisar o INCTCC, capturamos os dados dos pesquisadores (total = 122), sem introduzir na análise os alunos bolsistas envolvidos nos subprojetos.

A média de relacionamentos da rede é de **8,496**. A média de relacionamento da subrede dos nós mais conectados da rede é de **2,667**, sendo que 8 nós são pessoas sem nenhuma conexão com os demais pesquisadores do INCTCC, embora possuam ligações externas. Na rede há 8 pesquisadores que funcionam como articulações em nossa rede, ou seja, conectam grandes grupos. A detecção de articulações é importante para que possamos verificar os pontos de fragilidade de nossa rede, tornando-a frágil e facilmente dividida em subgrupos. Foram identificados 14 clusters através da métrica de *modularity*, porém desses, apenas 3 puderam ser associados a áreas do INCTCC. São eles: Medicina (20,51% dos pesquisadores), Saúde Coletiva (16,24%) e Genética (9,4%).

Desde a criação do INCT, em Janeiro de 2009, sendo a implementação oficial em junho do mesmo ano, pudemos notar modificações na rede. Principalmente se compararmos ela com o período pré-criação do INCT. Analisando apenas o relacionamento de co-autoria, verificamos que o grau médio desses anos foi: 1.009 (2007); 0,957 (2008); 0,410 (2009); 0,855 (2010) e 0,171 (2011). Nota-se que o número de interações dos especialistas estava decrescendo até o período da implantação do INCT em 2009. A diferença positiva (aumento de relacionamentos) entre os anos 2009-2010, ou seja, o primeiro ano de funcionamento do INCTCC foi de 0,445, superior à diferença negativa (diminuição de relacionamentos) do ano anterior (2008-2009, crescimento de -0,547), mostrando que as interações voltaram a crescer após a sua criação (Figura 3). Porém em 2011 houve um queda abrupta desse valor. É possível que isso tenha ocorrido pelo fato de muitas publicações ainda não estarem cadastradas no Lattes, até por questões de pendência de aprovação em *Journals*.



**Figura 3. Total de relacionamentos de co-autoria ao longo dos anos**



**Figura 4. Número de Publicações ao longo dos anos**

Seguindo essa mesma lógica de análise, notamos que novas relações surgiram de 2008 até 2011, mostrando que um dos principais objetivos do INCTCC estava sendo alcançado. Isso é algo muito positivo, pois mostra que mesmo o número total de relacionamentos tendo diminuído, o número de novas relações se manteve quase constante chegando a aumentar em 2010. Os dados que temos indica que em 2008

foram 54, 47 em 2009, 79 em 2010 e 25 em 2011. Mais uma vez devemos lembrar que 2011 ainda não está fechado, pois dependemos em grande parte de dados do Lattes, que são cadastrados pelos próprios pesquisadores.

Ainda pudemos analisar o número de publicações ao longo dos anos, como visto na Figura 4. Há uma queda de 2008 para 2009 de 12%, porém em 2010 a queda em comparação com 2008 é de 29%. Uma análise para isso seria que após a criação do INCTCC os pesquisadores que o integram passaram a publicar apenas artigos mais focados nas pesquisas que conduzem no INCT e menos em outras áreas de pesquisas paralelas. Outra possível explicação é pelo crescimento de novos relacionamentos. É natural que, ao começar a interagir com novas pessoas, exista um tempo de adaptação. Este processo de adaptação envolve o aprendizado sobre a pesquisa de seu novo parceiro, adequação a novos processos e métodos, além do amadurecimento da pesquisa e obtenção de novos resultados. A adaptação consome tempo e é natural que durante este período, não se tenham resultados a serem publicados.

#### **4. Conclusão e Trabalhos Futuros**

Este projeto ainda está em desenvolvimento e com ele podemos citar alguns trabalhos e melhorias futuras. O primeiro desafio é, com os dados coletados, realizar análises mais aprofundadas no cenário de pesquisa do câncer e cruzá-los com os dados de prontuários médicos e de tratamento do câncer. Outro obstáculo a ser superado refere-se à adoção de técnicas de mineração de dados para detectar regras de associação, ou seja, padrões de recorrência. Como último trabalho, iremos estudar os benefícios da abordagem criada para o cenário da pesquisa no controle do câncer no Brasil.

**Agradecimentos** ao Instituto Nacional de Ciência e Tecnologia para o Controle do Câncer, ao CNPq pelo apoio dado a projetos (476130/2009-3- Universal e 308219/2010-4, PQ2) e bolsas de estudos e CAPES por bolsas de estudos.

#### **Referências**

- Fernandes, G. O., Oliveira, J., Souza, J. M. (2011) “XMLattes A Tool for Importing and Exporting Curricula Data” In: IKE'11 - International Conference on Information and Knowledge Engineering, 2011, Las Vegas.
- Gephi (2012) “Gephi” <http://gephi.org/>, Abril.
- INCT (2012) “inct” [http://memoria.cnpq.br/programas/inct/\\_apresentacao/](http://memoria.cnpq.br/programas/inct/_apresentacao/), Abril.
- INCTCC (2010) “INCT Activity Report 2010 – home - INCA” [http://www1.inca.gov.br/inca/Arquivos/INCT/inct\\_project\\_2010.pdf](http://www1.inca.gov.br/inca/Arquivos/INCT/inct_project_2010.pdf), Abril.
- Lattes (2012) “Lattes” <http://lattes.cnpq.br/>, Abril.
- Pentaho (2012) “Pentaho Kettle Project”, <http://kettle.pentaho.com>, Abril.
- Pubmed (2012) “PubMed home”, <http://www.ncbi.nlm.nih.gov/pubmed>, Abril.
- Tableau (2012) “Tableau Software”, <http://www.tableausoftware.com/>, Abril.